DOCUMENT RESUME

ED 262 683 HE 018 752

AUTHOR Frierson, Henry T., Jr.
TITLE Assessing Intervention Effects on Repeated Standardized Test Performance: Examining Raw Percentage Scores.
PUB DATE [84]
NOTE 13p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Failure; Educational Testing; Higher Education; Intervention; *Medical Students; Response Style (Tests); *Scores; Skill Development; *Standardized Tests; Student Evaluation; *Test Wiseness
IDENTIFIERS *University of North Carolina

ABSTRACT

Effects of test-taking instruction for University of North Carolina medical students who failed end-of-year examinations were studied. The exam is compensatory and includes a number of subtests, including National Board (NB) Part I standardized subtests in pathology, pharmacology, and physiology. Students who had to repeat NB Part I subtests were assessed over a 5-year period. Two groups were evaluated: an intervention group of 15 students who participated in a special intervention program involving test-taking skills instruction; and a comparison group of 13 nonparticipants. Test-taking instruction consisted of: a systematic approach for taking tests, focusing on key words in both item stems and options, avoiding mistakes and increasing accuracy levels, analyzing perceived difficult items, and using guessing strategies as a last resort. The subtests were used as units of analyses, and the intervention group students retook 39 subtests compared to 33 for comparison group students. The intervention group's mean retest performance was significantly greater than that of the comparison group. It was concluded that intervention results in higher test performance and the implications are significant for medical students who have inadequate test-taking skills or for students undergoing academic remediation. (Author/SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED262683

ASSESSING INTERVENTION EFFECTS ON REPEATED STANDARDIZED TEST PERFORMANCE: EXAMINING RAW PERCENTAGE SCORES

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Henry T. Frierson, Jr.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

NE 0102052

Henry T. Frierson, Jr.
Office of Research and Development for
Education in the Health Professions
The University of North Carolina
School of Medicine
322 MacNider Building, 202H
Chapel Hill, N. C. 27514
U.S.A.
(919) 966-3641

ASSESSING INTERVENTION EFFECTS ON REPEATED STANDARDIZED TEST PERFORMANCE: EXAMING RAW PERCENTAGE SCORES

SUMMARY

Associated effects of test-taking skills related instruction over a five year period were assessed by examining retest performances of second year medical students who, because of unsatisfactory performance, were required to repeat National Board Part I subtests. The students were in two groups: the Intervention Group had 15 students who participated in a special intervention program involving test-taking skills instruction; the Comparison Group included 13 nonparticipants. The subtests were used as units of analyses, and the students in the Intervention Group retook 39 subtests compared to 33 for those in the Comparison Group. The Intervention Group's mean retest performance was significantly greater ($p < .001$) than that of the Comparison Group, and intervention effect size was 1.18. The results indicated that the intervention was associated with higher test performance and the implicatons are significant for medical students demonstrating low performances on standardized type exams because of inadequate test-taking skills or for students undergoing academic remediation.

INTRODUCTION

The average medical student faces a plethora of academic exams that must be passed for purposes of grades and promotion. But warranting cause for concern are factors unrelated to subject mastery that can depress test performance.

Those factors can be sources of significant measurement error attributing to substantial variation in test scores (Stanley, 1971). Cognitive factors such as the lack of test-wiseness have been cited as examples of error sources associated with lowered test performance (Millman, Bishop, & Ebel, 1965; Wahlstrom & Boerma, 1968; Sarnacki, 1979), as have noncognitive factors such as anxiety and stress (Sarason, 1975). Further, Nunnally (1978) reported that carelessness and confusion by test-takers are significantly related to measurement error and accompanying lowered test performances. Other factors possibly related to depressed test scores are some test-takers' perceptions that exams are threatening and intimidating.

Concerning the effects of measurement error on performance, Nunnally (1978) reported that obtained scores are biased estimates of true scores: i.e., scores above the mean are biased upward, and scores below the mean are biased downward. The latter situation implies the presence of negative measurement error and suggests that for any group or individual scoring significantly below the overall mean of the population, the possibility of large measurement error exists. Since problems related to test-taking also occur for medical students, that concept certainly has implications for those students experiencing low performance on tests. For example, it is probable that comparatively poor test-taking skills may be an underlying cause for some medical students who experience difficulty with exams as indicated by performances trends considerably below class means.

Relatedly, both Ebel (1979) and Stanley (1971) represent those who suggest the need for developing students' test-taking skills so their test performances are more likely to reflect acquired knowledge. That notion certainly seems applicable to the medical shool setting, and a related approach was attempted in a special intervention program at the University of North Carolina School of

Medicine for students who had demonstrated unsatisfactory exam performances particularly on standardized tests. But in light of the high reliability of standardized tests, researchers such as Messick (1980; Messick & Jungeblut, 1981) questioned the effectiveness of attempts to improve standardized test scores through intervention. The purpose of this study, however, was to assess the effects of test-taking related instruction for students who had demonstrated failing performances on end-of-year exams. The following question was addressed: Is intervention related to test-taking skills instruction associated with higher retest scores on standardized board subtests for medical students?

BACKGROUND

At the University of North Carolina, second year medical students must pass an end-of-year comprehensive exam to be promoted to the third year. That exam is compensatory and includes a number of subtests of which three are National Board (NB) Part I standardized subtests in Pathology, Pharmacology, and Physiology. Each subtest has a specified weight and is scored separately. The scores are then combined to form a total comprehensive test score. Each NB subtest had a weight of 17 percent and hence a combined weight of 51 percent of the total exam. Students were required to achieve a total score at least one-and-one-half standard deviations below the class mean. Subtests were examined separately to uncover any weak performances in specific subjects. Hence, students failing to achieve a passing score on the comprehensive exam did so because of unsatisfactory performance on specific subtests, and the majority of those performed unsatisfactorily on one or more of the heavily weighted NB subtests. Instead of repeating the academic year, most students failing the comprehensive exam are allowed to retake failed subtests after participating in

Medicine for students who had demonstrated unsatisfactory exam performances particularly on standardized tests. But in light of the high reliability of standardized tests, researchers such as Messick (1980; Messick & Jungeblut, 1981) questioned the effectiveness of attempts to improve standardized test scores through intervention. The purpose of this study, however, was to assess the effects of test-taking related instruction for students who had demonstrated failing performances on end-of-year exams. The following question was addressed: Is intervention related to test-taking skills instruction associated with higher retest scores on standardized board subtests for medical students?

BACKGROUND

At the University of North Carolina, second year medical students must pass an end-of-year comprehensive exam to be promoted to the third year. That exam is compensatory and includes a number of subtests of which three are National Board (NB) Part I standardized subtests in Pathology, Pharmacology, and Physiology. Each subtest has a specified weight and is scored separately. The scores are then combined to form a total comprehensive test score. Each NB subtest had a weight of 17 percent and hence a combined weight of 51 percent of the total exam. Students were required to achieve a total score at least one-and-one-half standard deviations below the class mean. Subtests were examined separately to uncover any weak performances in specific subjects. Hence, students failing to achieve a passing score on the comprehensive exam did so because of unsatisfactory performance on specific subtests, and the majority of those performed unsatisfactorily on one or more of the heavily weighted NB subtests. Instead of repeating the academic year, most students failing the comprehensive exam are allowed to retake failed subtests after participating in

a two month Summer Review Program (SRP); satisfactory retest performance is neccessary before they can be promoted to the third year.

During the SRP, the students receive tutorial assistance from faculty in areas of subject matter weakness indicated by exam performance. As most of the SRP students were required to retake one or more NB subtests, those who expressed concern about taking standardized exams, had the option of participating in a special intervention program that provided instruction in test-taking skills relative to NB type items. Over a period of time it was observed that those students who participated in the special intervention program produced superior retest performances.

METHODS

Study Participants

Over a five year period, 32 second year medical students repeated one or more of the three NB subtests. Of those students, 15 voluntarily participated in the special intervention program on a regular basis (at least two sessions per week over a minimum period of six weeks), 13 chose not to participate, and four participated only occasionally. The subtest scores of students in the first two groups were examined. The remaining four students were excluded since their number was too small to be considered as a third group.

Criterion Variable

Notably, all the students in the SRP did not take the same subtests or the same number of subtests, therefore the recorded scores of each NB subtest were used as units of analyses. Since there was such a variation of required retests, the combined performance on all subtests was employed as a dependent

variable for assessing effects associated with intervention. The 15 intervention participants retook a total of 39 subtests (2.60 per student), and the 13 nonparticipants retook 33 subtests (2.54 per student). Subtest scores were reported as percent correct, and the reliability coefficients (KR-20) of the subtests were generally in the 0.80 to 0.85 range.

Intervention Procedure

Instruction related to test-taking included at least two sessions per week. Each session was one to two hours in length. The sessions were conducted on both group and individual levels; this was dependent on the number of participating students per subject area. For example, in situations where one student was required to retake one exam in a subject not required for others, that student would participate in intervention sessions on an individual basis.

The objective of the first week was to introduce effective test-taking techniques that decreased the occurrence of unwarranted errors. During that period, test-taking problems experienced by the students on National Board type items were discussed. The students were then introduced to effective test-taking approaches. The test-taking instruction included the following components:

- A general systematic approach for taking tests
- Focusing on key words in both item stems and options
- Avoiding mistakes and increasing accuracy levels
- Analysis of perceived difficult items
- The use of guessing strategies as a last resort.

Basic test-taking instructions essentially comprised the first week.

The overall objective of the second week was to have the students apply the methods introduced in the first week. Using practice test, the students went

through the steps related to the suggested test-taking approach. They responded to items similar to those found on National Board Exams and then would receive immediate feedback on both their test-taking methods and selected answers.

The general objective for the third week was to allow the students to become more accustomed to using the suggested test-taking methods. The students were instructed to apply the suggested techniques on brief (30-35 items) practice tests. Items on the practice test contained content from the basic science tutorial portion of the SRP. Thus, the practice exams covered material generally perceived as relevant by the students. To further simulate actual test conditions, beginning in the second half of the week, a time limit was established for completing the exams.

Since simulated test contained content the students studied or were studying for the tutorial portion of the SRP, the use of simulated tests for assessing levels of knowledge was also emphasized in the third week. The students were encouraged to use the simulated tests to also help determine strengths and weaknesses in content knowledge.

All subsequent sessions, which varied in number, were a continuation of the third week: Effective test-taking practices were reinforced and self-assessment procedures were emphasized. Most sessions after the initial three week period involved timed, simulated exams that were of varying lengths (30-100 items).

Analysis

T-tests were used to analyze retest score differences. Intervention effect size (mean difference divided by the Comparison Group's standard deviation) was examined, and the grops' gain scores were separately analyzed and compared. A test for homogeneity of variance showed the variances of the two groups' initial mean subtest scores were equivalent ($F = 1.12$, $\underline{df} = 70$). Further analysis

indicated the groups' initial NB subtest means were quite similar ($t = 0.02$, $df = 70$) as shown by the scores--51.77 (SD = 5.99) and 51.90 (SD = 5.37)--for the Intervention and Comparison groups, respectively.

Results

Both groups had significant gains on the retests. The Comparison Group showed an increase from 51.90 to 58.12 (S.D. = 4.66) ($t = 4.22$, $df = 32$, $p < .001$), and the Intervention Group showed an increase from 51.77 to 63.63 (S.D. = 6.68) ($t = 9.84$, $df = 38$, $p < .0001$). The Intervention Group's mean retest performance, however, was significantly higher than that of the nonparticipants ($t = 4.14$, $df = 70$, $p < .001$). Further, the effect size of intervention on retest scores at 1.18 indicated the degree of nonoverlap for the two groups' score distributions was significant: the average Intervention Group member's score was higher than 86 percent of those in the Comparison Group.

In further examining group performance differences, the School's five year mean standard devdiation of 9.23 and mean score of 68.25 was used as reference points. The Comparison Group showed a 0.67 (6.22) standard deviation gain whereas the Intervention Group attained a 1.28 (11.86) gain--0.61 standard deviation greater than the Comparison Group.

DISCUSSION

Although the average SRP participant improved his performance, the greatest improvement was attained by those who participated in the special intervention program. Hence, the results indicate that intervention in the form of test-taking instruction was associated with higher retest scores and thus did not support Messick's (1980; Messick & Jungeblut, 1981) argument regarding

standardized exams, that marked improvements via short-term intervention would be difficult to achieve. The results further showed that the Intervention Group's mean retest score was well within one standard deviation of the school mean. Consequently, another contention is that the Intervention Group's retest mean was within acceptable performance limits relative to the school mean.

The results thus have significant implications for students who demonstrate low test performances because of inadequate test-taking skills or for students--who may also have poor test-taking skills--undergoing remediation because of unacceptable academic performance. The improved performance of the Intervention Group points to the practical value of students acquiring effective test-taking skills that enable them to demonstrate acquired knowledge more accurately on important examinations. Since the use of tests for grading and promotion purposes will, in all likelihood, remain a fact of life in most medical schools, it would seem productive to forthrightly address the effects of test-taking problems on student performance.

The greater performance associated with the applied intervention points to the practical importance of students acquiring effective test-taking skills. The acquisition of those skills will help students experiencing test-taking difficulties to more accurately demonstrate knowledge; and thus reduce unwarranted errors on important curricular related tests as well as standardized licensing type examinations such as National Boards. Extensive efforts, therefore, should occur to assure that students are exposed to methods that can reduce potential test-taking problems and thus reduce the the level of unwarranted errors that can result in depressed test scores.

Because of methodological limitations such as the lack of randomization and variances in the number and type of repeated subtests taken, cautious interpretation of the results is warranted. Consequently, further

investigations in this area is recommended, and whenever possible, under more controlled conditions. However, the educational significance of the results of the current study is arguably quite clear, as are the implications, particularly when viewed relative to those students regularly affected by low test performances.

REFERENCES

1. Ebel, R.L. Essentials of Educational Measurement. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1979.

2. Messick, S. The effectiveness of coaching for the SAT: Review and reanalysis of research from the fifties to the FTC. Princeton, N.J.: Educational Testing Service, 1980.

3. Messick, S. & Jungeblut, A. Time and method in coaching for the SAT. Psychological Bulletin, 1981, 89, 191-216.

4. Millman, J., Bishop, C.H., & Ebel, R. An Analysis of Test-Wiseness. Educational and Psychological Measurements. 1965, 25, 707-726,.

5. Nunnally, J.C. Psychometric Theory. New York, NY: McGraw Hill, 1978.

6. Sarason, I.G. Anxiety and Self-Preoccupation. In Sarason, I.G. & Spielberger, C.D. (Eds.), Stress and Anxiety (Vol. 2). New York: John Wiley & Sons, 1975.

7. Sarnacki, R.E. An Examination of Test-Wiseness in the Cognitive Domain. Review of Educational Research. 1979, 49, 252-279.

8. Stanley, J.C. Reliability. In Thorndike, R.L. (ed): Educational Measurement (2nd ed). Washington, D.C.: American Education Council, 1971

9. Wahlstrom, M. & Boerma F. The Influence of Test-Wiseness Upon Achievement. Educational and Psychological Measurements. 1968, 28, 413-420.